

VU Research Portal

A coskewness shrinkage approach for estimating the skewness of linear combinations of random variables

Boudt, Kris; Cornilly, Dries; Verdonck, Tim

published in

Journal of Financial Econometrics
2020

DOI (link to publisher)

[10.1093/JJFINEC/NBY022](https://doi.org/10.1093/JJFINEC/NBY022)

document version

Publisher's PDF, also known as Version of record

document license

Article 25fa Dutch Copyright Act

[Link to publication in VU Research Portal](#)

citation for published version (APA)

Boudt, K., Cornilly, D., & Verdonck, T. (2020). A coskewness shrinkage approach for estimating the skewness of linear combinations of random variables. *Journal of Financial Econometrics*, 18(1), 1-23.
<https://doi.org/10.1093/JJFINEC/NBY022>

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

E-mail address:

vuresearchportal.ub@vu.nl

A Coskewness Shrinkage Approach for Estimating the Skewness of Linear Combinations of Random Variables*

Kris Boudt ^{1,2}, Dries Cornilly^{1,3} and Tim Verdonck³

¹Vrije Universiteit Brussel, ²Vrije Universiteit Amsterdam and ³KU Leuven

Address correspondence to Dries Cornilly, Vrije Universiteit Brussel, Pleinlaan 2, 1050 Bruxelles, Belgium; KU Leuven, Celestijnenlaan 200B, 3001 Leuven, Belgium, or e-mail: dries.cornilly@kuleuven.be

Received February 2, 2018; revised August 6, 2018; editorial decision August 10, 2018; accepted August 12, 2018

Abstract

Decision-making in finance often requires an accurate estimate of the coskewness matrix to optimize the allocation to random variables with asymmetric distributions. The classical sample estimator of the coskewness matrix performs poorly for small sample sizes. A solution is to use shrinkage estimators, defined as the convex combination between the sample coskewness matrix and a target matrix. We propose unbiased consistent estimators for the MSE loss function and include the possibility of having multiple target matrices. In a portfolio application, we find that the proposed shrinkage coskewness estimators are useful in mean–variance–skewness efficient portfolio allocation of funds of hedge funds.

Key words: coskewness, MSE, multiple targets, portfolio optimization, shrinkage

JEL classification: C100, C130, G110

Accurate estimation of the skewness of linear combinations of p random variables is a key concern in finance. The estimation and management of the third-order interactions between financial returns are becoming increasingly important in asset allocation, portfolio management and risk analysis. The traditional approach to solving the problem is to estimate the skewness directly from the sample values of that linear combination (Hosking, 1990). An alternative, which is especially popular in optimizing and analyzing financial portfolios, is to obtain the skewness statistic using an estimate of the $p \times p^2$ third-order comoment matrix measuring the third-order interactions between the random variables. The estimation

* We thank the Editor (Fabio Trojani), two anonymous referees, and participants at various conferences and seminars for their valuable comments. We are especially grateful to Christophe Croux, Emmanuel Jurczenko, Siem Jan Koopman, Wanbo Lu, and Brian G. Peterson for stimulating discussions.

of such a coskewness matrix faces a similar curse of dimensionality as the estimation of the covariance matrix.

Though the higher order moments are becoming increasingly important, the number of available estimation methods is limited compared to the burgeoning literature on covariance estimation. In the seminal paper by [Martellini and Ziemann \(2010\)](#), the work on linear shrinkage of covariance matrices by [Ledoit and Wolf \(2003, 2004\)](#) is generalized to the use of shrinkage for the estimation of the coskewness matrix. Their extension includes proposing an estimator for the optimal shrinkage intensity in terms of Mean Squared Error (MSE). However, the proposed estimator is not consistent and uses a biased estimator for the MSE loss function. Moreover, their approach is limited to shrinking towards a single target matrix. Recently, [Jondeau, Jurczenko, and Rockinger \(2017\)](#) propose moment component analysis as a higher order equivalent of principal component analysis.

In this article, we contribute to the literature on skewness shrinkage estimation in four ways. First, we propose unbiased consistent estimators for the MSE loss function, which leads to an improved estimation of the shrinkage intensity. Second, under the MSE loss function, we derive optimal targets for a given coskewness structure. Third, we extend the methodology to include multiple targets simultaneously, eliminating the need to choose a single target matrix. Fourth, we show on simulated and real-life return data that the computational complexity of estimating a $p \times p^2$ coskewness matrix pays off in terms of a more precise estimate of the skewness of a linear combination of skewed variables. This result is of direct interest for decision-making based on higher order approximations of the expected utility function (see e.g., [Jondeau and Rockinger, 2006](#)) and [Martellini and Ziemann, 2010](#)), the density function (see e.g., [Boudt, Peterson, and Croux, 2008](#), [Del Brio, Niguez, and Perote, 2009](#) and [Stoyanov, Rachev, and Fabozzi, 2013](#)), or the construction of mean-variance-skewness (MVS) efficient portfolios using the shortfall measure of [Briec, Kerstens, and Jokung \(2007\)](#).

We illustrate our methodology on hedge fund return data for which accurate estimates of the coskewness matrix are needed to construct MVS efficient portfolios in the framework of [Briec, Kerstens, and Jokung \(2007\)](#). Rolling estimation windows are used to account for the transient nature of skewness of financial returns ([Beedles, 1979](#); [Singleton and Wingender, 1986](#)). In our setting, there are more assets than observations, leading to a misspecified sample coskewness matrix and the explicit need for shrinkage. The out-of-sample evaluation of investment performance shows that the proposed MSE optimized shrinkage estimators not only solve the issue from a statistical perspective, but also lead to substantial economic gains for the fund of hedge funds investor with moment preferences ([Kane, 1982](#); [Scott and Horvath, 1980](#)).

The remainder of the article is organized as follows. Section 1 introduces the notation and the traditional coskewness estimators. Section 2 describes single-target linear shrinkage estimation of the coskewness matrix and extends this framework to include multiple targets. Section 3 provides both the plug-in estimators and our proposed unbiased consistent estimators. The good performance of the proposed estimators is illustrated in an extensive simulation study in Section 4. Finally, we document the usefulness of the shrinkage estimators for optimizing portfolios of hedge funds in Section 5. We end with a conclusion and some policy implications.

A [Supplementary Appendix](#) discusses the impact of time series dependence on the problem of estimating the coskewness matrix. It also contains additional simulation results and details for the empirical application. We show how to include several other coskewness matrices from the literature into the multi-target shrinkage framework introduced in this article and extend the in-sample study of [Martellini and Ziemann \(2010\)](#) by including all proposed and corrected estimators. Finally, we demonstrate the **R** code for our estimators, which is publicly available in the PerformanceAnalytics package of [Peterson and Carl \(2018\)](#).

1 Coskewness Estimation

In this section, we first introduce the notation for the remainder of the article. We then present the sample estimator and the structured estimation approach, which will serve as the basis of the shrinkage estimator proposed in Section 2.

1.1 Notation

Let $(\mathbf{x}_1, \dots, \mathbf{x}_n)$ with $\mathbf{x}_i \in \mathbb{R}^p$ be a sample of n independent and identically distributed p -dimensional vectors drawn from the distribution of a random variable \mathbf{X} with mean $\boldsymbol{\mu}$ and coskewness matrix Φ . The matrix Φ consists of the third-order central moments

$$\varphi_{ijk} = \mathbb{E}[(X_i - \mu_i)(X_j - \mu_j)(X_k - \mu_k)], \quad i, j, k = 1, \dots, p, \quad (1)$$

where X_i and μ_i are, respectively, the i th component of the random variable \mathbf{X} and the mean $\boldsymbol{\mu}$. The coskewness elements are stacked in the coskewness matrix as

$$\Phi = \mathbb{E}[(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})' \otimes (\mathbf{X} - \boldsymbol{\mu})'], \quad (2)$$

where \otimes denotes the Kronecker product. Equivalently, we have that the coskewness matrix can be written as

$$\Phi = (\Phi_1 | \dots | \Phi_p), \quad (3)$$

with

$$\Phi_k = \begin{pmatrix} \varphi_{k11} & \cdots & \varphi_{k1p} \\ \vdots & \ddots & \vdots \\ \varphi_{kp1} & \cdots & \varphi_{kpp} \end{pmatrix}. \quad (4)$$

Note that φ_{ijk} has the same value for each permutation of the indices i, j , and k . This property is called supersymmetry of the matrix Φ . The coskewness matrix Φ is of dimension $p \times p^2$, but due to the property of supersymmetry, only $p(p+1)(p+2)/6$ elements are unique.

For a linear combination $\mathbf{v}'\mathbf{X}$ of \mathbf{X} , define the univariate skewness φ_v as

$$\varphi_v = \mathbb{E}[(\mathbf{v}'\mathbf{X} - \mathbf{v}'\boldsymbol{\mu})^3] = \mathbf{v}'\Phi(\mathbf{v} \otimes \mathbf{v}). \quad (5)$$

This article uses boldface letters to denote matrices and vectors. For matrices \mathbf{A} and \mathbf{B} , the inner product is defined by $\langle \mathbf{A}, \mathbf{B} \rangle = \text{trace}(\mathbf{A}'\mathbf{B})$. The Frobenius norm is denoted by $\|\cdot\|$ and it holds that $\langle \mathbf{A}, \mathbf{A} \rangle = \|\mathbf{A}\|^2$. In addition, $\xrightarrow{a.s.}$ and \xrightarrow{p} denote convergence almost surely and convergence in probability, respectively, always for $n \rightarrow \infty$. When \rightarrow is used,

convergence of a sequence of real numbers is meant. Finally, $f(n) = O(g(n))$ if and only if there exists $n_0 \in \mathbb{N}$ and $M \in \mathbb{R}^+$ such that $|f(n)| \leq M|g(n)|$ for all $n \geq n_0$.

1.2 Sample Estimator

Probably, the most intuitive way to estimate Φ is by the plug-in method, where expectations are replaced by sample averages. First, the mean μ is estimated by the sample average defined as

$$\bar{x} = \frac{1}{n} \sum_{l=1}^n x_l. \quad (6)$$

By replacing the mean in (1) with the sample mean and the expectation with a sample average, we obtain the plug-in estimator used in [Martellini and Ziemann \(2010\)](#) for a coskewness element

$$\hat{\varphi}_{ijk}^{\text{pl}} = \frac{1}{n} \sum_{l=1}^n (x_{li} - \bar{x}_i)(x_{lj} - \bar{x}_j)(x_{lk} - \bar{x}_k). \quad (7)$$

The above estimator is biased. As an alternative, we recommend to use the unbiased estimator for φ_{ijk} , see for example, [Fisher \(1929\)](#), given by

$$\hat{\varphi}_{ijk} = \frac{n}{(n-1)(n-2)} \sum_{l=1}^n (x_{li} - \bar{x}_i)(x_{lj} - \bar{x}_j)(x_{lk} - \bar{x}_k). \quad (8)$$

Note that the constant $n/((n-1)(n-2))$ is in essence a small sample correction, with a larger impact than the factor $1/(n-1)$ for unbiased covariance estimation.

Throughout the paper, we denote by $\hat{\Phi}$ the sample coskewness estimator stacking the individual elements estimated in [Equation \(8\)](#) as

$$\hat{\Phi} = \frac{n}{(n-1)(n-2)} \sum_{l=1}^n (x_l - \bar{x})(x_l - \bar{x})' \otimes (x_l - \bar{x})'. \quad (9)$$

We remark that this matrix format is mathematically convenient, but numerically inefficient. For optimal computational speed and memory use, exploiting the symmetry of the coskewness matrix is advised.

In [Appendix A](#), we use classical convergence results (see e.g., [Serfling 2009](#)) to prove that the sample estimator $\hat{\varphi}_{ijk}$ converges almost surely to the true coskewness value when $n \rightarrow \infty$. Proofs for all other properties in this article are also given in [Appendix A](#).

Property 1.1 *Assume that X has finite third-order moments. Then for any sample size n , $\mathbb{E}[\hat{\varphi}_{ijk}] = \varphi_{ijk}$, and when $n \rightarrow \infty$, it holds that $\hat{\varphi}_{ijk} \xrightarrow{a.s.} \varphi_{ijk}$.*

The unbiased sample estimator of the univariate skewness of $\nu'X$ in [Equation \(5\)](#) equals

$$\hat{\varphi}_\nu = \frac{n}{(n-1)(n-2)} \sum_{l=1}^n (\nu'x_l - \nu'\bar{x})^3. \quad (10)$$

Computing the skewness of a linear combination of random variables by using the univariate sample skewness is equivalent to first computing the sample coskewness estimator and then transforming the result to obtain the univariate estimate.

Property 1.2 For any $\mathbf{v} \in \mathbb{R}^p$, it holds that

$$\hat{\phi}_{\mathbf{v}} = \mathbf{v}'\hat{\Phi}(\mathbf{v} \otimes \mathbf{v}). \quad (11)$$

When $p > n$, there are infinitely many vectors \mathbf{v} such that $\mathbf{v}'\mathbf{x}_i = c, i = 1, \dots, n$ and hence $\hat{\phi}_{\mathbf{v}} = 0$. In this case, we call the sample coskewness matrix misspecified. A similar observation is made by [Engle \(2009\)](#) in the case of the sample covariance matrix.

1.3 Structured Coskewness Estimation

The sample estimator $\hat{\Phi}$ becomes unreliable in high dimensions. It attempts to estimate $p(p+1)(p+2)/6$ coskewness parameters using n observations and is even misspecified when $p > n$. This curse of dimensionality results in a low estimation precision, which can be avoided by regularizing the coskewness estimator or imposing some restrictions on its structure. In particular, when sample coskewness elements are similar, it may be possible to reduce the MSE by replacing the corresponding sample estimates by the sample average of these elements. This is a similar intuition as for the diagonal covariance matrix in [Ledoit and Wolf \(2004\)](#) or the block diagonal structure in [Devijver and Gallopin \(2017\)](#) and has a positive effect even if the restrictions are wrong. For $p = 2$, the structured coskewness matrix based on the independence and equal marginal assumption of [Ledoit and Wolf \(2004\)](#), can be written as follows:

$$\hat{T}_{LW}^* = \frac{\hat{\phi}_{111} + \hat{\phi}_{222}}{2} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \quad (12)$$

Note that in this notation, we use a supersymmetric matrix with only entries zero or one. This notation may seem daunting at first, but it is useful to obtain unbiased estimates of the MSE loss function when shrinking the sample coskewness matrix towards such target matrices, as will become clear in Section 2.1, in particular Property 2.2. In this notation, \hat{T}_{LW}^* belongs to the family of structured coskewness matrices outlined below.

Formally, let $E_q, q = 1, \dots, Q$ be supersymmetric matrices of dimension $p \times p^2$ containing only entries zero or one such that $\langle E_r, E_s \rangle = 0$ when $r \neq s$, that is, the matrices E_q do not have common 1-entries. Based on this, we propose to use the coskewness estimator \hat{T}^* defined by

$$\hat{T}^* = \sum_{q=1}^Q \frac{\langle E_q, \hat{\Phi} \rangle}{\|E_q\|^2} E_q. \quad (13)$$

We remark that the coefficients in \hat{T}^* are sample averages of the sample estimates $\hat{\phi}_{ijk}$ corresponding to the positions where $E_{q,ijk} = 1$. When $Q = p(p+1)(p+2)/6$, this corresponds to assuming that each coskewness element is unique. Hence, \hat{T}^* equals the sample coskewness matrix $\hat{\Phi}$. A low value of Q means that a lot of different coskewness elements are assumed to be roughly equal or zero. This might introduce an estimation bias, but possibly reduces the estimation variance drastically.

The structured coskewness matrices in Equation (13) are unbiased and consistent for the coskewness matrix T^* defined by

$$T^* = \sum_{q=1}^Q \frac{\langle E_q, \Phi \rangle}{\|E_q\|^2} E_q. \quad (14)$$

Property 1.3 Assume that X has finite third-order moments. For any sample size it holds that $\mathbb{E}[\hat{T}^*] = T^*$ and as $n \rightarrow \infty$, $\hat{T}^* \rightarrow^{a.s.} T^*$.

We proceed by giving three examples of structured coskewness matrices derived under this approach. The matrices are numbered because they will also be used in the simulation study in Section 4 and the empirical application in Section 5.

The simplest structured coskewness matrix is the deterministic matrix of dimension $p \times p^2$ containing only zeros, denoted by T_1 . This is the coskewness matrix arising from a random variable $X \in \mathbb{R}^p$ that is central-symmetric about a certain $\theta \in \mathbb{R}^p$. In particular, this holds for any elliptical distribution.

Coskewness matrices \hat{T}_2^* and \hat{T}_3^* assume independence and the structured coskewness matrix \hat{T}_2^* also assumes that the marginals have a common third-order central moment. They are given by

$$\hat{T}_2^* = \frac{\sum_{l=1}^p \hat{\varphi}_{lll}}{p} E_2, \quad \text{and} \quad \hat{T}_3^* = \sum_{q=1}^p \hat{\varphi}_{qqq} E_{3,q}, \quad (15)$$

with E_2 the $p \times p^2$ matrix for which all elements are equal to zero, except for the marginal third-order central moments, that is, the positions of φ_{lll} , $l = 1, \dots, p$ have as entry the value 1 and $E_{3,q}$, $q = 1, \dots, p$ are the matrices with the only non-zero element at position φ_{qqq} .

In the [Supplementary Appendix](#), we discuss other structured coskewness matrices available in the literature, namely the coskewness matrix under the latent single-factor model of [Simaan \(1993\)](#), the single-factor and constant correlation coskewness matrix of [Martellini and Ziemann \(2010\)](#) and the coskewness matrix under the multi-factor model of [Boudt, Lu, and Peeters \(2015\)](#).

2 Shrinkage Estimation

In this section, we present the single-target shrinkage estimator as a convex combination which combines the sample coskewness estimator $\hat{\Phi}$ and an alternative estimator \hat{T} into a shrinkage estimator of Φ . More formally, the single target shrinkage coskewness estimator is defined as

$$\hat{\Phi}^{\text{ST}}(\lambda) = (1 - \lambda)\hat{\Phi} + \lambda\hat{T}, \quad (16)$$

where the shrinkage intensity $\lambda \in [0, 1]$. The sample estimator $\hat{\Phi}$ of the coskewness matrix is an unbiased estimator, but typically has a large estimation variance depending on moments up to the sixth order of the distribution. The target matrix \hat{T} can be any $p \times p^2$ supersymmetric matrix for which $\hat{T} \neq \hat{\Phi}$. Typical choices for \hat{T} are the structured estimators in Equation (13), the coskewness matrix based on latent single-factor model ([Simaan, 1993](#)), the constant-correlation coskewness matrix or coskewness matrix based on an

observed single-factor model (Martellini and Ziemann, 2010) or the multi-factor coskewness matrix of Boudt, Lu, and Peeters (2015). The estimator \hat{T} of a structured coskewness matrix T usually has a lower estimation variance, but possibly is not consistent for the true coskewness matrix. A natural improvement is to combine both estimates into a single estimate by minimizing the MSE.

Single-target shrinkage for the coskewness matrix was proposed in Martellini and Ziemann (2010) for two particular structured coskewness matrices. We generalize their framework by introducing other (multiple) target matrices, proposing an unbiased and consistent estimator of the MSE loss function, and deriving the theoretical properties of the proposed estimators.

2.1 Optimization of the Shrinkage Intensity

In Martellini and Ziemann (2010), it is recommended to optimize the shrinkage intensity parameter λ in terms of the MSE loss function, given by

$$\text{which is equivalent to} \quad L(\lambda) = \mathbb{E}[\|(1 - \lambda)\hat{\Phi} + \lambda\hat{T} - \Phi\|^2], \quad (17)$$

$$L(\lambda) = A\lambda^2 - 2b\lambda + \mathbb{E}[\|\hat{\Phi} - \Phi\|^2], \quad (18)$$

where

$$A = \mathbb{E}[\|\hat{T} - \hat{\Phi}\|^2] \quad \text{and} \quad b = \mathbb{E}[\langle \hat{\Phi} - \hat{T}, \hat{\Phi} - \Phi \rangle]. \quad (19)$$

Clearly, the approach of optimizing the MSE only makes sense if the MSE in Equation (17) is finite. We, therefore, require that X has finite sixth order moments and the structured coskewness matrix \hat{T} has a finite MSE as well. Then the optimal shrinkage intensity, minimizing the MSE in Equation (17), is given by

$$\lambda^* = \frac{b}{A}, \quad (20)$$

which implicitly depends on the sample size through both A and b . We remark that, due to the unbiasedness of $\hat{\Phi}$, we can also write b as $\mathbb{E}[\|\hat{\Phi} - \Phi\|^2] - \mathbb{E}[\langle \hat{\Phi} - \Phi, \hat{T} - \mathbb{E}[\hat{T}] \rangle]$.

Given the optimal shrinkage intensity λ^* , the single-target shrinkage estimator $\hat{\Phi}^{\text{ST}}(\lambda^*)$ is consistent for Φ .

Property 2.1 Assume that X has finite sixth order moments and $\hat{T} \xrightarrow{p} U$ for some U such that $\text{Var}(\hat{t}_{ijk}) = O(n^{-1})$, then $\hat{\Phi}^{\text{ST}}(\lambda^*) \xrightarrow{p} \Phi$ as $n \rightarrow \infty$.

The second condition in Property 2.1 implies that $b = O(n^{-1})$. Hence, when $U \neq \Phi$, it holds that $\lambda = O(n^{-1})$. These conditions are weak. For target matrices \hat{T}^* in Equation (13), the conditions are satisfied when X has finite sixth order moments. Hence, for the target coskewness matrices used in this paper, the only condition is the existence of sixth order moments of X . The structured coskewness matrices in the Supplementary Appendix also satisfy the assumptions of Property 2.1 under mild assumptions on X .

Note that λ^* depends on the unknown A and b . Hence, the optimal shrinkage intensity needs to be estimated through estimation of A and b , which we propose in Section 4. If the shrinkage intensity λ^* is consistently estimated, it follows that the estimator $\hat{\Phi}^{\text{ST}}(\hat{\lambda}^*)$ is consistent as well.

By definition, the MSE of the shrinkage estimator is equal or lower than the MSE of the sample estimator. This decrease comes at the cost of a bias, introduced by the target matrix. However, the shrinkage estimator optimally balances the trade-off between estimation variance and bias by selecting the shrinkage intensity resulting in the lowest MSE.

Optimal coskewness shrinkage estimation is not only a matter of balancing the sample estimator and the target, but also of selecting an appropriate target matrix. In this regard, the structured coskewness matrices in Equation (13) can be seen as the solution minimizing the MSE with a more generic structured coskewness matrix. For a given choice of structure matrices $E_q, q = 1, \dots, Q$, the coefficients in T^* are optimal in the sense that they minimize the MSE.

Property 2.2 Assume that X has finite sixth order moments and a structured coskewness matrix $T = \sum_{q=1}^Q \nu_q E_q$, with ν_q scalars and the matrices E_q satisfying the constraints as for Equation (13). Then for any $\lambda \in (0, 1]$, the values of ν_q minimizing the MSE,

$$L(\nu_1, \dots, \nu_q) = \mathbb{E} \left[\left\| (1 - \lambda) \hat{\Phi} + \lambda \sum_{q=1}^Q \nu_q E_q - \Phi \right\|^2 \right], \quad (21)$$

are $\nu_q = \langle E_q, \Phi \rangle / \|E_q\|^2$ and thus T^* in Equation (13) is optimal for the structure given by the E -matrices.

2.2 Extension to Multi-Target Shrinkage

Often, several plausible calibrations of the target coskewness matrix \hat{T} exist. We then recommend to use shrinkage estimation with multiple targets. As for the multi-target shrinkage covariance estimator of Bartz, Höhne, and Müller (2014) and Lancewicki and Aladjem (2014), the target weights can be estimated in a data-driven manner by minimizing the MSE. This section introduces the multi-target shrinkage estimator for the coskewness matrix.

The proposed multi-target shrinkage coskewness estimator is given by

$$\hat{\Phi}^{\text{MT}}(\lambda) = \left(1 - \sum_{m=1}^t \lambda_m \right) \hat{\Phi} + \sum_{m=1}^t \lambda_m \hat{T}_m, \quad (22)$$

where t is the number of targets and \hat{T}_m is the target matrices ($m = 1, \dots, t$) of dimension $p \times p^2$. We require that none of the target matrices equals the sample estimator and that $\hat{\Phi} - \hat{T}_1, \dots, \hat{\Phi} - \hat{T}_t$ are linearly independent. In addition, we constrain the shrinkage intensities to be positive with sum less than unity,

$$\sum_{m=1}^t \lambda_m \leq 1 \quad \text{and} \quad \lambda_m \geq 0, \quad m=1, \dots, t. \quad (23)$$

In case $\hat{\Phi} - \hat{T}_1, \dots, \hat{\Phi} - \hat{T}_t$ are perfectly linearly dependent, it is possible to remove target matrices without loss of information or estimation accuracy. However, this condition is weak and holds with probability one for the structured coskewness matrices in this article and the Supplementary Appendix.

As in the single-target shrinkage case, we seek to find λ such that the MSE loss function

$$L(\lambda) = \mathbb{E} \left[\left\| \hat{\Phi}^{\text{MT}}(\lambda) - \Phi \right\|^2 \right] \quad (24)$$

is minimal, under the linear constraints in Equation (23). This optimization problem is a linearly constrained convex quadratic program. To show this, we first rewrite the loss function as

$$L(\lambda) = \lambda' A \lambda - 2b' \lambda + \mathbb{E}[\|\hat{\Phi} - \Phi\|^2], \quad (25)$$

where $A \in \mathbb{R}^{t \times t}$, $b \in \mathbb{R}^t$ are defined as

$$A_{ij} = \mathbb{E}[\langle \hat{T}_i - \hat{\Phi}, \hat{T}_j - \hat{\Phi} \rangle] \quad \text{and} \quad b_i = V(\hat{\Phi}) - C(\hat{\Phi}, \hat{T}_i), \quad i, j=1, \dots, t, \quad (26)$$

with

$$V(\hat{\Phi}) = \mathbb{E}[\|\hat{\Phi} - \Phi\|^2] \quad \text{and} \quad C(\hat{\Phi}, \hat{T}_i) = \mathbb{E}[\langle \hat{\Phi} - \Phi, \hat{T}_i - \mathbb{E}[\hat{T}_i] \rangle]. \quad (27)$$

Note that the matrix A is a Gramian matrix and hence is positive definite if and only if $\hat{\Phi} - \hat{T}_1, \dots, \hat{\Phi} - \hat{T}_t$ are linearly independent. Thus, for sufficiently different target matrices, the quadratic program is strictly convex with linear constraints and yields the unique solution λ^* . The convex nature of the problem guarantees finding the global optimum using standard quadratic solvers. The vector of shrinkage intensities nests as a special case the single-target shrinkage intensity in Equation (20).

The consistency result of the shrinkage estimator also holds in the multi-target setting.

Property 2.3 *Under the same conditions as Property 2.1 (for each of the targets) and the assumption that $\hat{\Phi} - \hat{T}_1, \dots, \hat{\Phi} - \hat{T}_t$ are linearly independent, $\hat{\Phi}^{\text{MT}}(\lambda^*) \xrightarrow{p} \Phi$ as $n \rightarrow \infty$.*

It should be noted that the optimal shrinkage intensity, with respect to the MSE, depends on the quantities b and A , which are unknown in practice.

3 Estimation of the Shrinkage Intensity

The optimal shrinkage coefficient λ^* that minimizes the MSE loss function $L(\lambda)$ in Equation (25), subject to the constraints Equation (23), depends on the unknown quantities A and b in Equation (26). An estimate of the optimal shrinkage intensity is obtained by minimizing the quadratic function

$$\hat{L}(\lambda) = \lambda' \hat{A} \lambda - 2\hat{b}' \lambda, \quad (28)$$

subject to the constraints Equation (23). The natural approach to estimate the matrix A is to use its sample version, that is,

$$\hat{A}_{ij} = \langle \hat{T}_i - \hat{\Phi}, \hat{T}_j - \hat{\Phi} \rangle, \quad i, j = 1, \dots, p. \quad (29)$$

Before providing the properties of this estimator, we note that for A_{ij} in Equation (26),

$$A_{ij} \rightarrow \langle T_i - \Phi, T_j - \Phi \rangle, \quad (30)$$

when $n \rightarrow \infty$. Define the limit matrix to be A_0 .

Property 3.1 *Under the assumptions of Property 2.3, it holds that $\mathbb{E}[\hat{A}_{ij}] = A_{ij}$ at any sample size n , and when $n \rightarrow \infty$, $\hat{A}_{ij} \xrightarrow{p} A_{0,ij}$.*

The estimation of \mathbf{b} is more complicated since $V(\hat{\Phi})$ and $C(\hat{\Phi}, \hat{T}_m)$ cannot simply be replaced by their sample realizations. An entry b_m of \mathbf{b} can be written as

$$b_m = \sum_{i,j,k=1}^p \left(\text{Var}(\hat{\varphi}_{ijk}) - \text{Cov}(\hat{\varphi}_{ijk}, \hat{t}_{m,ijk}) \right), \quad (31)$$

and hence, under the assumptions of Property 2.1 on \hat{T}_m , $n\mathbf{b} \rightarrow \mathbf{c}_0$ as $n \rightarrow \infty$, where

$$c_{0,m} = \sum_{i,j,k=1}^p \left(\text{AVar}(\sqrt{n}\hat{\varphi}_{ijk}) - \text{ACov}(\sqrt{n}\hat{\varphi}_{ijk}, \sqrt{n}\hat{t}_{m,ijk}) \right), \quad (32)$$

with AVar and ACov denoting the asymptotic variance and asymptotic covariance. Hence, $b_m = O(n^{-1})$ and a consistent estimator for \mathbf{c}_0 is required instead of any sequence of order $O(n^{-1})$ to estimate b_m .

In Section 3.1, we first describe the traditional approach of using plug-in estimators. Since this approach yields biased and non-consistent estimators for \mathbf{c}_0 , we recommend as an alternative to use estimators based on k -statistics and polykays, which we propose in Section 3.2.

3.1 Plug-in Estimation of \mathbf{b}

Plug-in estimation of the required quantities is the standard in the shrinkage literature. The principle is to replace each expectation by a sample average. An intuitive estimator for $V(\hat{\Phi})$ is given in [Martellini and Ziemann \(2010\)](#):

$$\hat{V}^{\text{pl}}(\hat{\Phi}) = \frac{1}{n^2} \sum_{l=1}^n \|\hat{\Phi}_l - \hat{\Phi}^{\text{pl}}\|^2, \quad (33)$$

where $\hat{\Phi}_l, l = 1, \dots, n$, is defined by

$$\hat{\Phi}_l = (\mathbf{x}_l - \bar{\mathbf{x}})(\mathbf{x}_l - \bar{\mathbf{x}})' \otimes (\mathbf{x}_l - \bar{\mathbf{x}})'. \quad (34)$$

This estimator is intuitive because it resembles the sample variance estimator.

Property 3.2 *When X has finite sixth order moments, it holds that*

$$\lim_{n \rightarrow \infty} nV(\hat{\Phi}) = \sum_{i,j,k=1}^p \text{AVar}(\sqrt{n}\hat{\varphi}_{ijk}), \quad i, j, k = 1, \dots, p, \quad (35)$$

where

$$\begin{aligned} \text{AVar}(\sqrt{n}\hat{\varphi}_{ijk}) = & m_{2,2,2} - m_{1,1,1}^2 - 2m_{2,1,1}m_{0,1,1} - 2m_{1,2,1}m_{1,0,1} - 2m_{1,1,2}m_{1,1,0} \\ & + m_{2,0,0}m_{0,1,1}^2 + m_{0,2,0}m_{1,0,1}^2 + m_{0,0,2}m_{1,1,0}^2 + 6m_{0,1,1}m_{1,0,1}m_{1,1,0}, \end{aligned} \quad (36)$$

and $m_{u,v,w}$ denotes the central moment, $m_{u,v,w} = \mathbb{E}[(X_i - \mu_i)^u (X_j - \mu_j)^v (X_k - \mu_k)^w]$.

Since $n\hat{V}^{\text{pl}}(\hat{\Phi}) \xrightarrow{a.s.} m_{2,2,2} - m_{1,1,1}^2$, as $n \rightarrow \infty$ and the extra terms in [Equation \(36\)](#) are not zero in general, the plug-in estimator is not consistent.

For a target \widehat{T}^* as in Equation (13), it can be shown that

$$\mathbb{E}[\langle \widehat{\Phi} - \Phi, \widehat{T}^* - \mathbb{E}[\widehat{T}^*] \rangle] = \mathbb{E}[\|\widehat{T}^* - \mathbb{E}[\widehat{T}^*]\|^2] = V(\widehat{T}^*). \quad (37)$$

The corresponding plug-in estimator, obtained by replacing expectations with sample averages, is given by

$$\widehat{V}^{\text{pl}}(\widehat{T}^*) = \frac{1}{n^2} \sum_{l=1}^n \|\widehat{T}_l^* - \frac{(n-1)(n-2)}{n^2} \widehat{T}^*\|^2, \quad (38)$$

where \widehat{T}_l^* is given by

$$\widehat{T}_l^* = \sum_{q=1}^Q \frac{\langle E_q, \widehat{\Phi}_l \rangle}{\|E_q\|^2} E_q. \quad (39)$$

Combining estimators (33) and (38) yields an estimate for b_m whenever the m -th target is of form Equation (13).

3.2 Unbiased Estimation of b

3.2.1 Estimation of $V(\widehat{\Phi})$

Unbiased estimators for both $V(\widehat{\Phi})$ and $C(\widehat{\Phi}, \widehat{T}_m^*)$ are obtained by element-wise estimation of $\text{Var}(\widehat{\varphi}_{ijk})$ and $\text{Cov}(\widehat{\varphi}_{ijk}, \widehat{t}_{m,ijk})$ as in Equation (31). For $\widehat{\varphi}_{iii}$, $i = 1, \dots, p$, the variance of the estimator is given by

$$\text{Var}(\widehat{\varphi}_{iii}) = \frac{1}{n} \kappa_6 + \frac{1}{n-1} \left(9\kappa_4\kappa_2 + 9\kappa_3^2 + \frac{6n}{n-2} \kappa_2^3 \right), \quad (40)$$

where κ_m denotes the m th cumulant of the distribution of X_i . For ease of notation, we omit the reference to i . Each of the terms in the right-hand side of Equation (40) is estimated by the corresponding unbiased estimator, as presented in Di Nardo, Guarino, and Senato (2008, 2009). Combining k -statistics and polykays, the unbiased estimator for Equation (40) is

$$\widehat{\text{Var}}(\widehat{\varphi}_{iii}) = c_{1,1}S_6 + c_{1,2}S_4S_2 + c_{1,3}S_3^2 + c_{1,4}S_2^3, \quad \text{where} \quad S_m = \sum_{l=1}^n (x_{li} - \bar{x}_i)^m. \quad (41)$$

Again, the reference to component i in S_m is omitted for ease of notation. The constants $c_{1,1}, \dots, c_{1,4}$ can be derived from the formula for the estimator $\widehat{\text{Var}}(\widehat{\varphi}_{ijk})$, presented in Appendix B. The resulting unbiased estimator for $V(\widehat{\Phi})$ is

$$\widehat{V}(\widehat{\Phi}) = \sum_{i,j,k=1}^p \widehat{\text{Var}}(\widehat{\varphi}_{ijk}). \quad (42)$$

In Section 4, we confirm that this estimator is an improvement over the biased and inconsistent plug-in estimator given in Section 3.1.

3.2.2 Estimation of $C(\widehat{\Phi}, \widehat{T}^*)$

Estimation of $C(\widehat{\Phi}, \widehat{T})$ has to be considered for each target coskewness matrix individually. In this section, we provide unbiased estimators when the target is as in Equation (13). The Supplementary Appendix contains consistent estimators for the other structured

coskewness matrices mentioned in Section 1.3. There we also correct the estimators given in [Martellini and Ziemann \(2010\)](#) for the single-factor and constant correlation coskewness matrices.

For any target \hat{T}^* , an unbiased estimator for $C(\hat{\Phi}, \hat{T}^*)$ can be constructed using multivariate k -statistics and polykays. Since T_1^* is deterministic, $C(\hat{\Phi}, T_1^*) = 0$. For target \hat{T}_2^* , with common third-order central moments it holds that

$$C(\hat{\Phi}, \hat{T}_2^*) = \frac{1}{p} \left(\sum_{i=1}^p \text{Var}(\hat{\varphi}_{iii}) + \sum_{i=1}^p \sum_{j \neq i}^p \text{Cov}(\hat{\varphi}_{iii}, \hat{\varphi}_{jjj}) \right) \quad (43)$$

and for \hat{T}_3^* the expression is

$$C(\hat{\Phi}, \hat{T}_3^*) = \sum_{i=1}^p \text{Var}(\hat{\varphi}_{iii}). \quad (44)$$

An estimator for the terms $\text{Cov}(\hat{\varphi}_{iii}, \hat{\varphi}_{jjj})$ using k -statistics and polykays is provided in Appendix B.

Combining the estimators $\widehat{\text{Var}}(\hat{\varphi}_{ijk})$ and $\widehat{\text{Cov}}(\hat{\varphi}_{ijk}, \hat{t}_{m,ijk})$, an unbiased and consistent estimator \hat{b} is obtained.

Property 3.3 For target matrices as in [Equation \(13\)](#) it holds that $n\hat{b} \xrightarrow{a.s.} c_0$ as $n \rightarrow \infty$ and, at any sample size n , it holds that $\mathbb{E}[\hat{b}] = b$.

Hence, due to Properties 3.1 and 3.3, it holds that the MSE loss function determining $\hat{\lambda}^*$ is consistently estimated. Thus, $\hat{\Phi}^{\text{MT}}(\hat{\lambda}^*) \xrightarrow{p} \Phi$, as $n \rightarrow \infty$, which includes the single-target shrinkage estimator as a special case.

4 Simulation Study

4.1 Set-Up

The simulation set-up is chosen such that the simulated data shows the same characteristics for variance, skewness, and kurtosis as observed in the dataset with monthly hedge fund returns used in Section 5. In a similar fashion as [Fan, Fan, and Lv \(2008\)](#), we estimate the multi-factor model

$$X = BF + \epsilon \quad (45)$$

on the 100 funds of strategies Equity Hedge, Macro, Relative Value, and Event-Driven according to Hedge Fund Research with most Assets Under Management (AUM) in December 2010. As factors, the corresponding HFRI indices (available on www.hedgefundresearch.com) are used.

As in [Jondeau and Rockinger \(2012\)](#), the factors are first standardized and then fitted by a skew- t distribution with density

$$f(x; \nu, \xi) = \frac{2b}{\xi + \frac{1}{\xi}} \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\pi(\nu-2)}\Gamma\left(\frac{\nu}{2}\right)} \left(1 + \frac{\kappa^2}{\nu-2}\right)^{-\frac{\nu+1}{2}}, \quad (46)$$

where $\kappa = (\beta x + \alpha)\xi$ if $\beta x + \alpha < 0$ and $(\beta x + \alpha)/\xi$ if $\beta x + \alpha \geq 0$. The parameters α and β are such that the distribution has mean zero and unit variance, namely

Table 1 PRIAL values in percentage for univariate skewness estimates using the multivariate shrinkage estimators

	<i>n</i>	<i>p</i> = 5		<i>p</i> = 10		<i>p</i> = 20		<i>p</i> = 30		<i>p</i> = 50		<i>p</i> = 100	
		Pl.	Unb.	Pl.	Unb.	Pl.	Unb.	Pl.	Unb.	Pl.	Unb.	Pl.	Unb.
$\widehat{\Phi}^{ST1}$	10	74.9	97.3	74.6	97.4	73.8	97.4	73.9	97.7	73.7	97.7	73.6	97.6
	20	88.9	94.8	88.2	94.3	87.7	94.2	88.0	94.5	87.6	94.2	87.2	93.8
	30	83.5	84.5	83.9	85.5	82.8	85.4	83.5	85.8	83.6	86.3	83.5	86.4
	50	83.1	84.5	82.9	84.7	80.8	83.3	80.8	83.3	80.9	83.5	80.8	83.5
	100	72.6	76.4	71.8	76.0	69.2	74.4	68.8	74.1	70.2	75.0	70.1	75.0
	250	47.5	56.5	47.2	55.9	46.6	56.0	44.1	54.5	50.5	58.1	52.5	59.6
	500	36.3	45.0	36.2	44.4	36.8	45.2	32.0	41.8	41.1	47.7	42.9	48.9
	1000	45.5	50.1	46.0	50.3	49.8	53.8	45.9	50.9	52.1	55.1	54.5	57.2
$\widehat{\Phi}^{ST2}$	10	68.2	95.8	72.8	97.3	73.4	97.4	73.6	97.7	73.6	97.7	73.6	97.6
	20	84.3	93.7	87.0	94.3	87.4	94.2	87.8	94.4	87.5	94.2	87.2	93.8
	30	81.2	83.4	83.3	85.5	82.8	85.4	83.2	85.7	83.6	86.3	83.5	86.4
	50	82.3	83.7	82.8	84.7	80.8	83.3	80.6	83.1	80.9	83.5	80.8	83.5
	100	74.0	76.0	72.3	76.0	69.4	74.4	68.5	73.8	70.2	74.9	70.1	75.0
	250	51.2	56.6	48.3	56.0	46.9	56.0	43.7	54.1	50.5	58.1	52.5	59.6
	500	38.9	44.5	37.0	44.3	37.0	45.2	31.6	41.4	41.1	47.7	42.9	48.9
	1000	45.1	48.1	46.0	49.9	49.8	53.7	45.6	50.6	52.1	55.1	54.5	57.2
$\widehat{\Phi}^{ST3}$	10	67.9	95.7	72.8	97.3	73.4	97.4	73.5	97.7	73.6	97.7	73.6	97.6
	20	84.0	93.7	87.0	94.3	87.4	94.2	87.8	94.4	87.5	94.2	87.2	93.8
	30	81.0	83.3	83.4	85.5	82.8	85.4	83.4	85.7	83.6	86.3	83.5	86.4
	50	82.3	83.6	82.9	84.7	80.9	83.4	80.8	83.0	80.9	83.5	80.8	83.5
	100	74.6	76.1	72.7	76.1	69.6	74.5	69.0	73.4	70.3	74.9	70.1	75.1
	250	53.4	57.0	49.2	56.2	47.3	56.1	43.7	53.2	50.6	58.2	52.5	59.6
	500	41.3	45.0	37.9	44.5	37.4	45.4	31.8	41.1	41.4	47.8	42.9	48.9
	1000	46.4	48.3	46.5	50.0	50.0	53.8	46.5	51.2	52.3	55.2	54.5	57.2
$\widehat{\Phi}^{MT}$	10	74.9	97.4	74.6	97.4	73.8	97.4	73.9	97.7	73.7	97.7	73.6	97.6
	20	88.9	95.0	88.2	94.4	87.7	94.2	88.0	94.4	87.6	94.2	87.2	93.8
	30	83.7	84.8	83.9	85.5	82.8	85.4	83.4	85.7	83.6	86.3	83.5	86.4
	50	83.3	84.8	82.9	84.7	80.8	83.3	80.7	83.0	80.9	83.5	80.8	83.5
	100	72.9	76.7	71.8	76.0	69.2	74.4	68.4	73.3	70.2	74.9	70.1	75.0
	250	47.9	56.8	47.2	55.9	46.6	56.0	42.3	52.7	50.4	58.0	52.5	59.6
	500	36.6	45.4	36.2	44.4	36.8	45.2	30.4	40.4	41.1	47.7	42.9	48.9
	1000	45.6	50.2	46.0	50.4	49.8	53.8	45.4	50.4	52.1	55.1	54.5	57.2

Notes: PRIAL values with respect to the sample estimator of the skewness of the sum of **X** are given for the shrinkage estimators when the data are generated according to a factor model (Section 4.1). Three different single-target estimators are used in addition to the multi-target estimator that uses all three targets. PRIAL values for the plug-in (Pl.) and unbiased (Unb.) estimators are given for sample sizes ranging from *n* = 10 to 1000 and dimensions ranging from *p* = 5 to 100. The PRIAL values are reported in percentage points. The background color ranks the PRIAL values for each combination of *n* and *p*; the darker, the higher the PRIAL.

$$\alpha = \frac{\Gamma\left(\frac{\nu-1}{2}\right)\sqrt{\nu-2}}{\sqrt{\pi}\Gamma\left(\frac{\nu}{2}\right)}\left(\zeta + \frac{1}{\zeta}\right) \quad \text{and} \quad \beta = \sqrt{\zeta^2 + \frac{1}{\zeta^2} - 1 - \alpha^2}.$$

(47)

Moments up to order 6 exist when $\nu > 6$. Hence, when fitting the distribution to the factors, we use a lower bound of $\nu = 7$. Factor loadings and idiosyncratic terms are obtained

by least squares regression based on the 60 monthly observations from January 2006 to December 2010. Each of the idiosyncratic terms is also modeled by a skew- t distribution.

We consider the dimensions $p = 5, 10, 20, 30, 50, 100$ and sample sizes $n = 10, 20, 30, 50, 100, 250, 500, 1000$. For p smaller than 100, we subset the data generating process to include only the p funds with the largest AUM. For each sample size and dimension, 10,000 samples are generated.

4.2 Results

The focus of our simulation study is the estimation accuracy of the skewness of a linear combination of random variables. In our simulations, we take the sum ($\mathbf{v} = \mathbf{1}_p$). Denote by $\hat{\varphi}_v^{(\cdot)}$ the estimator of $\varphi_v^{(\cdot)}$ obtained by replacing in (5) the coskewness matrix Φ by an estimator $\hat{\Phi}_v^{(\cdot)}$. Due to Property 1.2, the multivariate sample estimator yields the same MSE when estimating the skewness of the linearly transformed variable. For the shrinkage estimators this equivalence does not hold.

In Table 1, we show the improvement of the shrinkage estimators $\varphi_v^{(\cdot)}$ over the sample estimator φ_v in terms of MSE. This is done using the percentage relative improvement in average loss (PRIAL) frequently used in the shrinkage literature, see for example, Ledoit and Wolf (2003). The PRIAL of an estimator $\hat{\varphi}_v^{(\cdot)}$ for φ_v , compared with the sample estimator $\hat{\varphi}_v$, is defined by

$$\text{PRIAL}(\hat{\varphi}_v^{(\cdot)}) = \left(\frac{\mathbb{E}[\|\hat{\varphi}_v - \varphi_v\|^2] - \mathbb{E}[\|\hat{\varphi}_v^{(\cdot)} - \varphi_v\|^2]}{\mathbb{E}[\|\hat{\varphi}_v - \varphi_v\|^2]} \right) \times 100\%. \quad (48)$$

Note that the PRIAL of the sample skewness estimator $\hat{\varphi}_v$ for φ_v is zero by definition and the PRIAL cannot exceed 100%. A negative value indicates a larger MSE of $\hat{\varphi}_v^{(\cdot)}$ compared with the MSE of the sample estimator $\hat{\varphi}_v$.

Table 1 shows remarkable improvements in the accuracy of the estimated skewness of a linear combination of random variables when multivariate shrinkage estimators are used instead of the sample estimator. The proposed shrinkage estimators outperform the plug-in versions consistently for all dimensions and sample sizes considered. The improvements are as large as 25 percentage points. Finite sample properties result in reliable estimates when the sample size is small. Because the sample estimator is consistent, it is important to correctly estimate the size of the shrinkage intensity for large sample sizes in order to have a good PRIAL. The proposed shrinkage estimators still offer over 50% reduction in MSE compared to the sample estimator at a sample size of $n = 1000$. PRIAL values measured on the full coskewness matrix instead of on the linear combination show similar results and are given in the Supplementary Appendix. We remark that when $p > n$, the sample coskewness matrix is misspecified, whereas the shrinkage estimators are not.

A key question is whether the multi-target shrinkage estimator offers additional benefits compared to single-target shrinkage. Table 1 shows that for $p = 5$ and $p = 10$ the PRIAL of the multi-target shrinkage estimator is higher than for each of the single-target shrinkage estimators, indicated by the darker shade of gray. For higher dimensions, the PRIAL is either the highest or within one percentage point of the highest PRIAL of the individual single-target estimators. This behavior is to be expected since only p out of p^3 coskewness

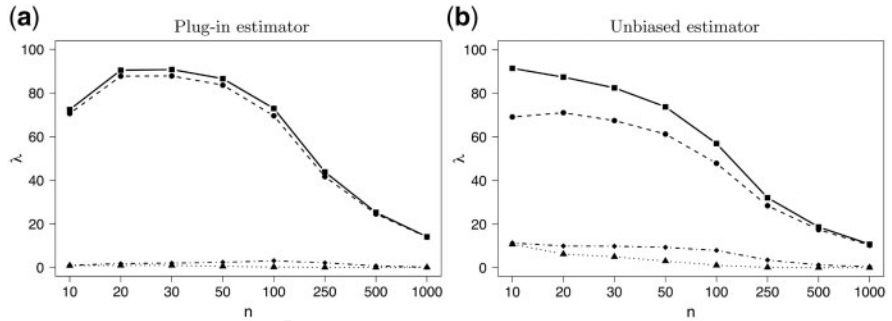


Figure 1 Mean shrinkage intensities of the multi-target estimators ($p = 50$). (a) Plug-in estimator (b) Unbiased estimator.

Note: The squares indicate the total shrinkage intensities of the multi-target estimator. Circles, triangles and diamonds denote, respectively, the shrinkage intensities for targets T_1^* , T_2^* and T_3^* .

elements is different between the targets. Hence, for larger dimensions, this results in almost identical targets. Even in this case, multi-target shrinkage offers advantages as it is able to select the “best” target. The best target is not necessarily the one closest to the true data generating process, but the one that offers the largest reduction in MSE when combined with the sample estimator. The gray shadings in Table 1 indicate that in this simulation setting, target T_1^* offers the largest reduction in MSE when the sample sizes are small. For larger sample sizes, \hat{T}_3^* becomes better. The results based on the full coskewness matrix are even more pronounced in showing the ability of the multi-target estimator to select the best target and realize a high PRIAL. In Section 5.4 of the Supplementary Appendix, we consider a simulation study where the data-generating process and thus the true coskewness matrix changes over the samples. Hence, different targets are expected to be the best depending on which process generated a particular sample. In this setting, the flexibility of the multi-target estimator really pays off and its PRIAL values are up to 40 percentage points higher than the best single-target shrinkage estimator.

A final finite sample properties question that we investigate through simulation is how the sample size affects the estimated multi-target shrinkage intensity parameters. To this end, we plot in Figure 1 the multi-target shrinkage intensity when $p = 50$ for a range of sample sizes, assuming the same data-generating process as in Table 1. The decomposition into the contribution of each target confirms the findings in PRIAL values that T_1^* provides a large reduction in MSE when the sample size is small. Note that its importance diminishes rather quickly when sample size increases. This is in contrast with \hat{T}_3^* for which the intensity decreases more slowly or even stays relatively constant, indicating that its relative importance in the contribution of the multi-target shrinkage estimator increases. Hence, \hat{T}_3^* becomes more informative when sample size increases. Also note the more natural decay of the shrinkage intensity of the proposed estimator which is in strong contrast with the hump shape seen under plug-in estimation.

In the Supplementary Appendix, we present simulation evidence confirming that our estimators for b are unbiased, and provide the PRIAL values computed for the full

coskewness estimators. These confirm the superiority of the proposed coskewness shrinkage estimators and validate the use of the multi-target estimator. In addition, we provide the bias and PRIAL results for two other simulation settings. First, the setting of an independent component model with Pareto distributed components and additive noise serves as a second example of a skewed and heavy-tailed distribution for which the targets are misspecified. Second, a Gaussian data-generating process is used to confirm that the estimators perform well even when X is symmetric.

5 Empirical Application

The above simulations have confirmed the statistical gains in accuracy when using the proposed shrinkage estimators. In this section, we further analyze the usefulness of these estimators for constructing MVS efficient fund of hedge funds portfolio. We follow the approach in [Briec, Kerstens, and Jokung \(2007\)](#) by seeking for the efficient portfolio that yields the highest improvements relative to a benchmark portfolio, which we set to the equally weighted (EW) portfolio. Denote this reference portfolio by w_0 . Suppose that the comoments up to order three are known. The resulting optimized portfolio is the solution to

$$\begin{aligned}
 & \underset{\delta \in \mathbb{R}, w_i \in [0,1]}{\text{maximize}} && \delta \\
 & \text{subject to} && w' \mu \geq w_0' \mu, \\
 & && w' \Sigma w \leq w_0' \Sigma w_0 (1 - \delta), \\
 & && w' \Phi(w \otimes w) \geq w_0' \Phi(w_0 \otimes w_0) (1 + s_0 \delta), \\
 & && \sum_{i=1}^p w_i = 1,
 \end{aligned} \tag{49}$$

where $s_0 = 1$ if $w_0' \Phi(w_0 \otimes w_0) \geq 0$ and $s_0 = -1$ otherwise. This means that when the optimal value δ^* is positive, there exists a portfolio with weights w^* such that its mean, variance, and skewness are more favorable compared to the initial portfolio. Formulation (49) implies a preference for a higher return, lower variance and due to s_0 , a higher skewness, which is the accepted direction of preference for the third-order central moment, see for example [Kane \(1982\)](#) or [Scott and Horvath \(1980\)](#). We remark that the dual approach with MVS approximation to an expected utility function, as in [Jondeau and Rockinger \(2006\)](#) or [Harvey et al. \(2010\)](#) does not guarantee a global optimum, while the primal approach in Equation (49) does, as proved in [Briec, Kerstens, and Jokung \(2007\)](#). The formulation in Equation (49) is infeasible as it assumes known moments. In practice, they need to be estimated, implying an estimation error in the optimized weights. We next evaluate the performance of the estimated portfolio for the proposed shrinkage estimators and compare it with benchmark allocations.

[Briec, Kerstens, and Jokung \(2007\)](#) illustrate their framework by computing the MVS efficient portfolio in a static example in a dimension of 35 assets. Our study compares dynamically optimized efficient portfolios to the EW portfolio, serving as a benchmark. For this, we consider a dataset of monthly returns of hedge funds belonging to the four main strategies (Equity Hedge, Macro, Relative Value, and Event-Driven), and having a history of at least 60 consecutive months over the period January 2000 until December 2013. At

Table 2. Out-of-sample performance of the portfolios

	EW	S	Single-target			Multi-target	
			T_1^*	\hat{T}_2^*	\hat{T}_3^*	$(\hat{T}_2^*, \hat{T}_3^*)$	$(T_1^*, \hat{T}_2^*, \hat{T}_3^*)$
Panel A: $p=40$							
Ann. Geom. Mean (10^{-2})	6.56	5.72	5.88	5.82	5.87	5.67	5.98
Ann. Standard Deviation (10^{-2})	5.82	3.55	3.48	3.63	3.40	3.59	3.52
Skewness (10^{-6})	-5.89	-1.50	-1.36	-1.37	-1.35	-1.37	-1.33
MUG over EW (bp., $\gamma = 15$)	0	92.98	112.71	99.32	115.00	86.93	119.83
MUG over EW (bp., $\gamma = 30$)	0	345.99	372.78	351.31	379.14	340.75	378.41
Break-even transaction costs (\$ $\gamma = 15$)		2.71	3.61	2.92	3.50	2.46	3.79
Break-even transaction costs (\$ $\gamma = 30$)		11.20	12.30	10.67	11.84	10.25	12.29
Mean shrinkage intensity (%)			85.33	85.30	85.34	86.38	86.42
Panel B: $p=100$							
Ann. Geom. Mean (10^{-2})	5.77	4.51	4.94	4.99	4.94	4.99	4.94
Ann. Standard Deviation (10^{-2})	6.16	3.74	3.78	3.78	3.77	3.78	3.78
Skewness (10^{-6})	-7.31	-1.92	-1.67	-1.68	-1.68	-1.68	1.67
MUG over EW (bp., $\gamma = 15$)	0	77.08	118.36	122.26	118.07	122.26	118.36
MUG over EW (bp., $\gamma = 30$)	0	381.73	426.98	430.61	426.94	430.61	426.98
Break-even transaction costs (\$ $\gamma = 15$)		2.32	5.70	5.50	5.56	5.50	5.70
Break-even transaction costs (\$ $\gamma = 30$)		12.27	19.38	19.06	18.84	19.06	19.38
Mean shrinkage intensity (%)			81.09	81.08	81.22	81.29	81.30

Notes: The table presents following out-of-sample performance measures for the EW and MVS efficient portfolios: annualized geometric mean and standard deviation, and skewness (non-standardized). In addition, we provide the annualized MUG for values of risk aversion $\gamma = 15$ and 30, the break-even transaction costs (dollar per \$1000 traded) for which the investor is indifferent between the EW and MVS portfolios, and the mean shrinkage intensity in the out-of-sample period. The EW portfolio is denoted by EW. We consider the following coskewness estimators: sample (S), single-target shrinkage estimators with targets T_1^* , \hat{T}_2^* and \hat{T}_3^* , and the multi-target shrinkage estimators with targets $(\hat{T}_2^*, \hat{T}_3^*)$ and $(T_1^*, \hat{T}_2^*, \hat{T}_3^*)$. The MVS portfolios are obtained by solving (49) and named after the respective coskewness estimator. Results are reported for an investment universe of $p = 40$ and 100 assets.

any point in time, the investment universe consists of the funds with most AUM, according to each strategy. We consider the dimensions 40 and 100, including the top 10 or top 25 funds in each strategy sorted by AUM.

To account for the potential time-variation in the coskewness matrix, we follow the industry practice of using 3-year rolling samples. In order to apply Equation (49), the expected return, covariance, and coskewness matrices are required at each time step. The expected return is estimated by the sample average and the covariance matrix by the shrinkage estimator of Ledoit and Wolf (2003) with a diagonal target. Details are discussed in the

Supplementary Appendix. To measure skewness, we consider the sample estimator, the single-target shrinkage estimators with targets T_1^* , \hat{T}_2^* and \hat{T}_3^* , and the multi-target shrinkage estimators with targets $(\hat{T}_2^*, \hat{T}_3^*)$ and $(T_1^*, \hat{T}_2^*, \hat{T}_3^*)$. In these settings, the sample coskewness matrix is ill-defined due to $p > n$ and shrinkage is necessary.

The performance is measured using out-of-sample realized moments. In addition to the annualized geometric mean and standard deviation, we provide the third-order central moment and mean shrinkage intensity over the out-of-sample period. Here, we present the results when a 3-year rolling window is used, while the **Supplementary Appendix** discusses the case of a 5-year rolling window. We measure economic value by means of the monetary utility gain (MUG) for investors with constant relative risk aversion (CRRA) (Ang and Bekaert, 2002; Martellini and Ziemann, 2010). In our application, the MUG equals the annual return in basis points required by investors in the EW portfolio so that they are indifferent to changing to the more complex MVS efficient investment strategy. To measure the relevance of such economic gains, we also report the break-even transaction costs, in dollars per \$1000 traded, for which a CRRA investor would be indifferent between the MVS portfolio and the EW portfolio.

Table 2 shows that in both dimensions, 40 and 100, the MVS efficient portfolios clearly realize a lower variance and a higher skewness. In case of the 40-fund investment universe, the multi-target shrinkage estimator with all three targets reduces the variance about 40% and increases the skewness by 75%. This is done at the cost of a one percentage point decrease in realized annual return.

In addition, the relative gain with respect to the portfolios based on the sample coskewness matrix is important. For both dimensions, we observe that the standard deviations are approximately equal across the optimized portfolios, but the shrinkage estimators achieve a higher skewness. Also, the multi-target shrinkage estimator with all three targets clearly improves over each of the single-target estimators individually.

The preferable moments indicate that the proposed estimator yields portfolios with a more attractive out-of-sample return distribution. Note that also the economic gains are substantial. In case of the 40-fund investment universe, the MUG values range between 86 and 378 basis points, depending on the risk aversion of the investor. This indicates that a CRRA investor clearly prefers the more sophisticated approach of investing in MVS efficient portfolios. The findings are robust over the different rolling window periods and coefficients of risk aversion, as studied in the **Supplementary Appendix**. We remark that the economic incentive to invest in MVS portfolios increases for increasing risk aversion. In addition, the break-even transaction costs range from \$2.4 to \$19.3 per \$1000 traded, indicating that even with transaction costs, there is an incentive for a CRRA investor to invest in the MVS portfolios instead of the EW one.

Overall, we may conclude that the multi-target coskewness shrinkage estimator achieves its goals of improving finite sample estimation of the coskewness matrix in general and regularizing the estimates when $p > n$ in particular, leading to an economic incentive to invest in MVS optimized portfolios constructed using shrinkage estimators.

6 Conclusion

Many financial decisions involve the evaluation or optimization of a higher order approximation of the expected utility function or the density of skewed random variables. The

quality of those decisions heavily depends on the accuracy of the estimates of the corresponding coskewness matrix.

The main message of this article is that, for the estimation of the skewness of linear combinations of random variables, one should consider a multivariate approach using a shrinkage estimate of the coskewness matrix, where the MSE loss function is estimated unbiasedly. Our simulations show that the unbiased estimators for the MSE loss function improve the finite sample behavior of the shrinkage estimator by up to 25 percentage points in terms of MSE. We further contribute by extending the methodology to accommodate the use of multiple targets and derive the optimal targets for a given coskewness structure.

In the empirical application, we show that there is an economic incentive for an investor with CRRA preferences to invest in MVS efficient portfolios constructed by estimating skewness using the proposed multi-target shrinkage coskewness estimator. This empirical evidence is in line with our simulation results about the gains in accuracy when estimating the skewness of linear combinations of random variables using a shrinkage-based estimate of the full coskewness matrix.

Our findings also have policy implications in terms of investor protection legislation. Many quantitative funds are still managed under the Markowitz mean–variance efficient allocation paradigm, and thus yield a suboptimal performance for the investor with skewness preferences. We hope that this paper and the associated implementation in the **R** package PerformanceAnalytics contribute to raising awareness of the importance of skewness among investors and policy makers.

Supplementary Data

Supplementary data are available at *Journal of Financial Econometrics* online.

Funding

This work was supported by the Research Foundation—Flanders [FWO research grant G023815N, PhD fellowship 1114117N to D.C.]; the Internal Funds KU Leuven (project C16/15/068); Google Summer of Code 2017; and the National Science Foundation of China [71771187 to K.B.].

Appendix A: Proofs

Proof of Property 1.1. Existence of third order moments ensures the existence of all variables in this proof. For the plug-in coskewness estimator $\hat{\phi}_{ijk}^{\text{pl}}$ it holds that

$$\begin{aligned} \hat{\phi}_{ijk}^{\text{pl}} &= \frac{1}{n} \sum_{l=1}^n (x_{li} - \mu_i)(x_{lj} - \mu_j)(x_{lk} - \mu_k) - (\bar{x}_i - \mu_i)\sigma_{jk} - (\bar{x}_j - \mu_j)\sigma_{ik} \\ &\quad - (\bar{x}_k - \mu_k)\sigma_{ij} - (\bar{x}_i - \mu_i)(\hat{\sigma}_{jk} - \sigma_{jk}) - (\bar{x}_j - \mu_j)(\hat{\sigma}_{ik} - \sigma_{ik}) \\ &\quad - (\bar{x}_k - \mu_k)(\hat{\sigma}_{ij} - \sigma_{ij}) + 2(\bar{x}_i - \mu_i)(\bar{x}_j - \mu_j)(\bar{x}_k - \mu_k), \end{aligned} \quad (50)$$

with $\sigma_{ij} = \text{Cov}(X_i, X_j)$ and $\hat{\sigma}_{ij} = n^{-1} \sum_{l=1}^n (x_{li} - \mu_i)(x_{lj} - \mu_j)$. By the Strong Law of Large Numbers and Application D of Corollary 1.7 in Serfling (2009), it follows that $\hat{\phi}_{ijk}^{\text{pl}} \xrightarrow{a.s.} \phi_{ijk}$, as $n \rightarrow \infty$. Since $n^2/((n-1)(n-2)) \rightarrow 1$ when $n \rightarrow \infty$, the same result holds for the

unbiased sample estimator. Taking the expected value of the right hand side of Equation (50), it holds that

$$\mathbb{E}[\hat{\phi}_{ijk}^{\text{pl}}] = \phi_{ijk} - \frac{3}{n} \phi_{ijk} + \frac{2}{n^2} \phi_{ijk} = \frac{(n-1)(n-2)}{n^2} \phi_{ijk}, \quad (51)$$

and hence the sample estimator $\hat{\phi}_{ijk}$ is unbiased.

Proof of Property 1.2. For any $\mathbf{v} \in \mathbb{R}^p$ we reformulate the expression of $\hat{\phi}_{\mathbf{v}}$ in (10) to

$$\begin{aligned} \hat{\phi}_{\mathbf{v}} &= \frac{n}{(n-1)(n-2)} \sum_{l=1}^n \left(\sum_{i=1}^p (v_i x_{li} - v_i \bar{x}_i) \right)^3 \\ &= \frac{n}{(n-1)(n-2)} \sum_{l=1}^n \sum_{i=1}^p \sum_{j=1}^p \sum_{k=1}^p (v_i x_{li} - v_i \bar{x}_i) (v_j x_{lj} - v_j \bar{x}_j) (v_k x_{lk} - v_k \bar{x}_k) \\ &= \sum_{i=1}^p \sum_{j=1}^p \sum_{k=1}^p v_i v_j v_k \hat{\phi}_{ijk} = \mathbf{v}' \hat{\Phi} (\mathbf{v} \otimes \mathbf{v}). \end{aligned} \quad (52)$$

Proof of Property 1.3. Observe that \hat{T}^* is a linear function of $\hat{\Phi}$. Hence, by Property (1.1) and Theorem 1.7 in Serfling (2009) it holds that $\hat{T}^* \xrightarrow{d.s.} T^*$ as $n \rightarrow \infty$ and $\mathbb{E}[\hat{T}^*] = T^*$.

Proof of Property 2.1. If $U = \Phi$, then for any sequence λ^* , $\hat{\Phi}^{\text{ST}}(\lambda) \xrightarrow{p} \Phi$ as $n \rightarrow \infty$. Assume that $U \neq \Phi$, then $A \rightarrow A_0 > 0$. By the Cauchy-Schwartz inequality, $b = O(n^{-1})$, and thus $\lambda^* = O(n^{-1})$. Hence, because $\hat{\Phi}$ is consistent, it holds that $\hat{\Phi}^{\text{ST}}(\lambda) \xrightarrow{p} \Phi$ as $n \rightarrow \infty$.

Proof of Property 2.2. The gradient of the MSE loss function (21) with respect to $(\nu_1, \dots, \nu_Q)'$ has entries

$$\frac{\partial}{\partial \nu_q} L = 2\lambda^2 \mathbb{E}[\langle E_q, \nu_q E_q - \hat{\Phi} \rangle], \quad q = 1, \dots, Q. \quad (53)$$

Solving the first order conditions and checking the second order conditions yields the optimal solution $\nu_q = \langle E_q, \Phi \rangle / \|E_q\|^2$, which are the coefficients in T^* .

Proof of Property 2.3. By definition of the quadratic program (25) and the restriction of λ^* to the convex set (23) it holds that

$$\mathbb{E}[\|\hat{\Phi}^{\text{MT}}(\lambda^*) - \Phi\|^2] \leq \min_{m=1, \dots, t} \mathbb{E}[\|\hat{\Phi}^{\text{ST}}(\lambda_m^*) - \Phi\|^2], \quad (54)$$

where $\hat{\Phi}^{\text{ST}}(\lambda_m^*)$ is the single-target shrinkage estimator with target \hat{T}_m and λ_m^* the optimal shrinkage intensity for this estimator. The proof of Property 2.1 implies that

$$\mathbb{E}[\|\hat{\Phi}^{\text{ST}}(\lambda_m^*) - \Phi\|^2] \rightarrow 0, \quad \text{as } n \rightarrow \infty. \quad (55)$$

Hence,

$$\mathbb{E}[\|\hat{\Phi}^{\text{MT}}(\lambda^*) - \Phi\|^2] \rightarrow 0, \quad \text{as } n \rightarrow \infty, \quad (56)$$

implying that $\hat{\Phi}^{\text{MT}}(\lambda^*) \xrightarrow{p} \Phi$ as $n \rightarrow \infty$.

Proof of Property 3.1. Due to Property 1.1 and the consistency assumption for the estimators of the target matrices, it follows that, by Theorem 1.7 and Application D of Corollary

1.7 in Serfling (2009), $\hat{A}_{ij} \xrightarrow{P} A_{0,ij}$, as $n \rightarrow \infty$. The estimator is unbiased by definition, since $\mathbb{E}[\hat{A}_{ij}] = \mathbb{E}[(\hat{T}_i - \hat{\Phi}, \hat{T}_j - \hat{\Phi})] = A_{ij}$.

Proof of Property 3.2. Multiplying Equation (50) with \sqrt{n} and applying Slutsky's lemma to the last four terms, it holds that

$$\begin{aligned} \sqrt{n}\hat{\phi}_{ijk}^{\text{pl}} &= \sqrt{n} \left(\frac{1}{n} \sum_{l=1}^n (x_{li} - \mu_i)(x_{lj} - \mu_j)(x_{lk} - \mu_k) \right. \\ &\quad \left. - (\bar{x}_i - \mu_i)\sigma_{jk} - (\bar{x}_j - \mu_j)\sigma_{ik} - (\bar{x}_k - \mu_k)\sigma_{ij} \right) + o_P(1), \end{aligned} \quad (57)$$

with $o_P(1)$ a term converging in probability to zero. The right hand side is a mean over independent and identically distributed observations with a variance that is equal to (36).

Proof of Property 3.3. The estimators $\widehat{\text{Var}}(\hat{\phi}_{ijk})$ and $\widehat{\text{Cov}}(\hat{\phi}_{iii}, \hat{\phi}_{jjj})$ needed to construct \hat{b} can be found in Appendix B. Since $n^{-1}S_{u,v,w} \xrightarrow{a.s.} m_{u,v,w}$ as $n \rightarrow \infty$, it follows that

$$n\widehat{\text{Var}}(\hat{\phi}_{ijk}) \xrightarrow{a.s.} \text{AVar}(\sqrt{n}\hat{\phi}_{ijk}), \quad \text{as } n \rightarrow \infty, \quad (58)$$

with $\text{AVar}(\sqrt{n}\hat{\phi}_{ijk})$ given in Equation (36). Analogously,

$$n\widehat{\text{Cov}}(\hat{\phi}_{iii}, \hat{\phi}_{jjj}) \xrightarrow{a.s.} \text{ACov}(\sqrt{n}\hat{\phi}_{iii}, \sqrt{n}\hat{\phi}_{jjj}), \quad \text{as } n \rightarrow \infty. \quad (59)$$

Hence, it holds that $n\hat{b} \xrightarrow{a.s.} c_0$ as $n \rightarrow \infty$ and by the properties of polykays and k -statistics that at any sample size n , $\mathbb{E}[\hat{b}] = b$.

Appendix B: Unbiased Estimators

Here we provide complete formulas for the $\widehat{\text{Var}}(\hat{\phi}_{ijk})$ and $\widehat{\text{Cov}}(\hat{\phi}_{iii}, \hat{\phi}_{jjj})$ needed to estimate nb consistently and unbiasedly. The formulas for the variances of the sample coskewness estimators in terms of multivariate cumulants can be found in Stuart and Ord (1994). The framework developed in Di Nardo, Guarino, and Senato (2008, 2009) is used to construct the unbiased estimators.

The variance of the sample coskewness

$$\begin{aligned} \text{Var}(\hat{\phi}_{ijk}) &= \frac{1}{n} \kappa_{2,2,2} + \frac{1}{n-1} \left(\kappa_{2,0,0}\kappa_{0,2,2} + \kappa_{0,2,0}\kappa_{2,0,2} + \kappa_{0,0,2}\kappa_{2,2,0} + 2\kappa_{1,1,0}\kappa_{1,1,2} \right. \\ &\quad \left. + 2\kappa_{1,0,1}\kappa_{1,2,1} + 2\kappa_{0,1,1}\kappa_{2,1,1} + 3\kappa_{1,1,1}^2 + 2\kappa_{2,1,0}\kappa_{0,1,2} + 2\kappa_{1,2,0}\kappa_{1,0,2} + 2\kappa_{2,0,1}\kappa_{0,2,1} \right) \\ &\quad + \frac{n}{n-2} \left(\kappa_{2,0,0}\kappa_{0,2,0}\kappa_{0,0,2} + \kappa_{2,0,0}\kappa_{0,1,1}^2 + \kappa_{1,1,0}\kappa_{0,0,2} + \kappa_{1,0,1}^2\kappa_{0,2,0} + 2\kappa_{1,1,0}\kappa_{0,1,1}\kappa_{1,0,1} \right), \end{aligned} \quad (60)$$

where $\kappa_{u,v,w}$ denotes the three-dimensional cumulant of order (u, v, w) of the random vector (X_i, X_j, X_k) . An unbiased estimator is given by

$$\begin{aligned} \widehat{\text{Var}}(\hat{\phi}_{ijk}) &= c_1 S_{2,2,2} + c_2 (S_{2,2,0}S_{0,0,2} + S_{2,0,2}S_{0,2,0} + S_{0,2,2}S_{2,0,0}) + c_3 (S_{2,1,1}S_{0,1,1} \\ &\quad + S_{1,2,1}S_{1,0,1} + S_{1,1,2}S_{1,1,0}) + c_4 (S_{0,2,1}S_{2,0,1} + S_{0,1,2}S_{2,1,0} + S_{1,0,2}S_{1,2,0}) + c_5 S_{1,1,1}^2 \\ &\quad + c_6 S_{0,0,2}S_{0,2,0}S_{2,0,0} + c_7 (S_{2,0,0}S_{0,1,1}^2 + S_{0,2,0}S_{0,1,1}^2 + S_{0,0,2}S_{1,1,0}^2) + c_8 S_{0,1,1}S_{1,0,1}S_{1,1,0}, \end{aligned} \quad (61)$$

where $S_{u,v,w} = \sum_{l=1}^n (x_{li} - \bar{x}_i)^u (x_{lj} - \bar{x}_j)^v (x_{lk} - \bar{x}_k)^w$ and

$c_1 = \alpha(n^6 - 5n^5 + 13n^4 - 23n^3 + 22n^2 - 8n)$, $c_2 = \alpha(-n^4 + 4n^3 - 9n^2 + 14n - 8)$,
 $c_3 = \alpha(-2n^5 + 12n^4 - 18n^3 - 16n^2 + 56n - 32)$, $c_4 = \alpha(-2n^4 + 8n^3 - 2n^2 - 20n + 16)$,
 $c_5 = \alpha(-n^5 + 2n^4 + 17n^3 - 34n^2 - 40n + 32)$, $c_6 = \alpha(4n^2 - 12n + 8)$,
 $c_7 = \alpha(n^4 - 8n^3 + 25n^2 - 34n + 16)$ and $c_8 = \alpha(6n^4 - 48n^3 + 134n^2 - 156n + 64)$,
 with $\alpha = (n(n-1)^2(n-2)^2(n-3)(n-4)(n-5))^{-1}$.

The covariance between two estimates of marginal skewness equals

$$\text{Cov}(\hat{\phi}_{iii}, \hat{\phi}_{jjj}) = \frac{1}{n} \kappa_{3,3} + \frac{1}{n-1} \left(9\kappa_{2,2}\kappa_{1,1} + 9\kappa_{1,2}\kappa_{2,1} + \frac{6n}{n-2} \kappa_{1,1}^3 \right), \quad (62)$$

for which an unbiased estimator is given by

$$\widehat{\text{Cov}}(\hat{\phi}_{iii}, \hat{\phi}_{jjj}) = c_1 S_{3,3} + c_9 S_{3,0} S_{0,3} + c_{10} S_{2,1} S_{1,2} + c_{11} S_{3,1} S_{0,2} + c_{11} S_{1,3} S_{2,0} + c_{12} S_{2,2} S_{1,1} + c_{13} S_{2,0} S_{0,2} S_{1,1} + c_{14} S_{1,1}^3, \quad (63)$$

where

$c_9 = \alpha(-n^5 + 5n^4 + 5n^3 - 31n^2 - 10n + 8)$, $c_{10} = \alpha(-9n^4 + 36n^3 - 9n^2 - 90n + 72)$,
 $c_{11} = \alpha(-3n^5 + 21n^4 - 39n^3 + 3n^2 + 42n - 24)$, $c_{12} = \alpha(-9n^4 + 36n^3 - 81n^2 + 126n - 72)$,
 $c_{13} = \alpha(9n^4 - 72n^3 + 189n^2 - 198n + 72)$ and $c_{14} = \alpha(24n^2 - 72n + 48)$.

References

- Ang, A., and G. Bekaert. 2002. International Asset Allocation with Regime Shifts. *Review of Financial Studies* 15: 1137–1187.
- Bartz, D., J. Höhne, and K.-R. Müller. 2014. Multi-Target Shrinkage. *arXiv Preprint arXiv* 1412: 2041.
- Beedles, W. L. 1979. On the Asymmetry of Market Returns. *Journal of Financial and Quantitative Analysis* 14: 653–660.
- Boudt, K., W. Lu, and B. Peeters. 2015. Higher Order Comoments of Multifactor Models and Asset Allocation. *Finance Research Letters* 13: 225–233.
- Boudt, K., B. G. Peterson, and C. Croux. 2008. Estimation and Decomposition of Downside Risk for Portfolios with Non-Normal Returns. *The Journal of Risk* 11: 79–103.
- Briec, W., K. Kerstens, and O. Jokung. 2007. Mean-Variance-Skewness Portfolio Performance Gauging: A General Shortage Function and Dual Approach. *Management Science* 53: 135–149.
- Del Brio, E. B., T.-M. Níguez, and J. Perote. 2009. Gram–Charlier Densities: A Multivariate Approach. *Quantitative Finance* 9: 855–868.
- Devijver, E., and M. Gallopin. 2017. Block-Diagonal Covariance Selection for High-Dimensional Gaussian Graphical Models. *Journal of the American Statistical Association* 113: 113–314.
- Di Nardo, E., G. Guarino, and D. Senato. 2008. A Unifying Framework for k-Statistics, Polykays and Their Multivariate Generalizations. *Bernoulli* 14: 440–468.
- Di Nardo, E., G. Guarino, and D. Senato. 2009. A New Method for Fast Computing Unbiased Estimators of Cumulants. *Statistics and Computing* 19: 155–165.
- Engle, R. 2009. *Anticipating Correlations: A New Paradigm for Risk Management*. Princeton and Oxford: Princeton University Press.
- Fan, J., Y. Fan, and J. Lv. 2008. High Dimensional Covariance Matrix Estimation Using a Factor Model. *Journal of Econometrics* 147: 186–197.
- Fisher, R. A. 1929. Moments and Product Moments of Sampling Distributions. *Proceedings of the London Mathematical Society* 30: 199–238.

- Harvey, C. R., J. C. Liechty, M. W. Liechty, and P. Müller. 2010. Portfolio Selection with Higher Moments. *Quantitative Finance* 10: 469–485.
- Hosking, J. R. 1990. L-Moments: Analysis and Estimation of Distributions Using Linear Combinations of Order Statistics. *Journal of the Royal Statistical Society. Series B (Methodological)* 52: 105–124.
- Jondeau, E., E. Jurczenko, and M. Rockinger. 2017. Moment Component Analysis: An Illustration with International Stock Markets. *Journal of Business & Economic Statistics* (Forthcoming).
- Jondeau, E., and M. Rockinger. 2006. Optimal Portfolio Allocation under Higher Moments. *European Financial Management* 12: 29–55.
- Jondeau, E., and M. Rockinger. 2012. On the Importance of Time Variability in Higher Moments for Asset Allocation. *Journal of Financial Econometrics* 10: 84–123.
- Kane, A. 1982. Skewness Preference and Portfolio Choice. *Journal of Financial and Quantitative Analysis* 17: 15–25.
- Lancewicki, T., and M. Aladjem. 2014. Multi-Target Shrinkage Estimation for Covariance Matrices. *IEEE Transactions on Signal Processing* 62: 6380–6390.
- Ledoit, O., and M. Wolf. 2003. Improved Estimation of the Covariance Matrix of Stock Returns with an Application to Portfolio Selection. *Journal of Empirical Finance* 10: 603–621.
- Ledoit, O., and M. Wolf. 2004. A Well-Conditioned Estimator for Large-Dimensional Covariance Matrices. *Journal of Multivariate Analysis* 88: 365–411.
- Martellini, L., and V. Ziemann. 2010. Improved Estimates of Higher-Order Comoments and Implications for Portfolio Selection. *Review of Financial Studies* 23: 1467–1502.
- Peterson, B. G., and P. Carl. 2018. *PerformanceAnalytics: Econometric Tools for Performance and Risk Analysis*. R Package Version 1.5.2.
- Scott, R. C., and P. A. Horvath. 1980. On the Direction of Preference for Moments of Higher Order than the Variance. *The Journal of Finance* 35: 915–919.
- Serfling, R. J. 2009. *Approximation Theorems of Mathematical Statistics*, Volume 162. New York: John Wiley & Sons.
- Simaan, Y. 1993. Portfolio Selection and Asset Pricing-Three Parameter Framework. *Management Science* 39: 568–577.
- Singleton, J. C., and J. Wingender. 1986. Skewness Persistence in Common Stock Returns. *Journal of Financial and Quantitative Analysis* 21: 335–341.
- Stoyanov, S. V., S. T. Rachev, and F. J. Fabozzi. 2013. Sensitivity of Portfolio VaR and CVaR to Portfolio Return Characteristics. *Annals of Operations Research* 205: 169–187.
- Stuart, A., and J. K. Ord. 1994. *Kendall's Advanced Theory of Statistics*. London: Edward Arnold.